UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/684,773	10/14/2003	Zicheng Liu	M61.12-0547	3356

27366 7590 01/12/2006

WESTMAN CHAMPLIN (MICROSOFT CORPORATION)
SUITE 1400 - INTERNATIONAL CENTRE
900 SECOND AVENUE SOUTH
MINNEAPOLIS, MN 55402-3319

EXAMINER	
WOODS, ERIC V	

ART UNIT	PAPER NUMBER
2672	

DATE MAILED: 01/12/2006

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)	
	10/684,773	LIU ET AL.	
	Examiner	Art Unit	
	Eric Woods	2672	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *30 September 2005*.

2a)☒ This action is **FINAL**. 2b)☐ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-43* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-43* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____.
5) ☐ Notice of Informal Patent Application (PTO-152)
6) ☐ Other: _____.

DETAILED ACTION

Specification

The title is accepted.

Response to Arguments

Applicant's arguments with respect to claims 1-43 have been considered but are moot in view of the new ground(s) of rejection, which are necessitated by applicant's amendments to the claims.

The objection to the title stands withdrawn in view of applicant's arguments on page 1 of Remarks submitted 30 September 2005.

All objections to the claims have been obviated by the amendments to the claims submitted 30 September 2005, and/or adequately traversed on pages 1-2 of Remarks submitted 30 September 2005.

First of all the amendment to include the limitation of a single frame is meaningless because a video consists of many frames so a system that synthesizes video will prima facie synthesize a single frame, and in any case the canons of claims construction interpret singular usage (as in the articles 'a' or 'an') as encompassing the plural (see Scanner Technologies Corp. v. ICOS Vision Systems Corp., 70 USPQ2d 1900 (CA FC 2004)). Therefore, a plurality of single frames constitutes video. Therefore, this portion of the amended claims is moot vis-à-vis any reference that generates video rather than a 'single frame'.

Claim Rejections - 35 USC § 103

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

The factual inquiries set forth in *Graham* **v.** *John Deere Co.*, 383 U.S. 1, 148 USPQ 459 (1966), that are applied for establishing a background for determining obviousness under 35 U.S.C. 103(a) are summarized as follows:

1. Determining the scope and contents of the prior art.
2. Ascertaining the differences between the prior art and the claims at issue.
3. Resolving the level of ordinary skill in the pertinent art.
4. Considering objective evidence present in the application indicating obviousness or nonobviousness.

Claim 1 is rejected under 35 U.S.C. 103(a) as being unpatentable over Yonezawa (US PGPub 2001/0036860 A1) in view of Moulton et al (US PGPub 2002/0097380 A1).

As to claim 1,

A computer implemented method for rendering a single frame of a synthesized image, comprising: (Preamble is not given patentable weight, since it only recites a summary of the claim and/or an intended use, and the process steps and/or apparatus components are capable of standing on their own; see Rowe v. Dror, 112 F.3d 473, 42 USPQ2d 1550 (Fed. Cir. 1997), Pitney Bowes, Inc. v. Hewlett-Packard Co., 182 F.3d 1298, 1305, 51 USPQ2d 1161, 1165 (Fed. Cir. 1999), and the like.)

-Generating a geometric component corresponding to a selected image for the frame based on identified feature points from a set of representative images, where each

image of the set has the identified feature points, and wherein the geometric component is a dimensional vector of feature point positions; and (Yonezawa Figure 8 – the user identifies and selects various feature points on a plurality of images – see image 400, where the user identifies control points C1-C3 ('control point group') and then tracks them to an intermediate image, where the control points are then matched to those in another image 403 with control points B1-B5 ('control point group') which are approximated between images. These can be of a mouth, the edges of eyes, or the like. See [0027-0034], where Yonezawa teaches that the conventional morphing method uses the same number of feature points per frame, where the method of Yonezawa can also do so. Clearly, two images constitute 'a set of representative images', and each image can have the same number of feature points. Clearly, these feature points are tracked between sets. See for example Figures 5 and 6, where the same number of points can be tracked between frames.)(Moulton clearly teaches that vectors represent the control points of each image and their motion path in the set of archival video footage and that they are tracked between frames [0011, 0059-0061]. See Figure 1 additionally)

-Generating the selected image for the frame from a composite of the set of representative images based on the geometric component. (Yonezawa generates intermediate images as in Figure 8, the abstract, and in other locations [0034].)(Moulton clearly generates new video footage, as in Figure 1, blocks 310-380 additionally.)

The section in the Response to Arguments block above concerning the single frame limitation is incorporated by reference.

Yonezawa teaches most of the limitations of the above claim, but does not expressly teach that the geometric component is a dimensional vector of feature point positions. Moulton is a system that tracks the locations of feature points along a motion path, such that the transition from viseme and/or phoneme to another can be tracked, and teaches the tracking of feature points between frames as well. Such transitions also include emotional ones – note step 340, Figure 1 and [0011, 0015, 0019, 0027], and the like. If an individual or a character in the system of Yonezawa were speaking and/or changed emotions in a video game [0002-0003], the speech and emotions would need to be synchronized, as per the system of Moulton. The techniques of Moulton for tracking feature points between frames would be ideal for use in the system of Yonezawa, which is silent on how the feature points are tracked in the intermediately generated frames per se. However, keeping such data in vector format makes sense, because the plurality of intermediate data generated would be in video format, for example for a video game. Moulton teaches that this technique provides advantages [0020, 0043, 0045, 0057] by allowing the user to associate trajectory paths with transitions and use a standard model for this. Yonezawa utilizes three-dimensional models [0064], where perspective projection techniques are used to render three-dimensional models and scenes (such as a 3D model of a face) to the two-dimensional screen, as does Moulton (inherently, by generating video). Therefore, for at least the above reasons, it would have been obvious to one of ordinary skill in the art at the time the invention was made to modify Yonezawa to use vectors to store feature point paths and the like. Note that Yonezawa does not teach away from using the conventional

morphing technique, but rather merely extends it to allow use in situations where the same number of control points may not be available.

Claims 1-14 and 24-25 are rejected under 35 U.S.C. 103(a) as being unpatentable over Cosatto in view of Moulton et al as above.

As to claim 1,

A computer implemented method for rendering a single frame of a synthesized image, comprising: (Preamble is not given patentable weight, since it only recites a summary of the claim and/or an intended use, and the process steps and/or apparatus components are capable of standing on their own; see Rowe v. Dror, 112 F.3d 473, 42 USPQ2d 1550 (Fed. Cir. 1997), Pitney Bowes, Inc. v. Hewlett-Packard Co., 182 F.3d 1298, 1305, 51 USPQ2d 1161, 1165 (Fed. Cir. 1999), and the like.)

-Generating a geometric component corresponding to a selected image for the frame based on identified feature points from a set of representative images, where each image of the set has the identified feature points, and wherein the geometric component is a dimensional vector of feature point positions; and (Cosatto teaches in section 1, page 152, that in the first step, image samples of facial parts are generated and results in a database of facial parts. Pages 153-154, section III, teach the methods of how this is done, and how the hierarchy of parts and samples are obtained and subsequently ordered. Section IV on page 154 states that the first step in the process is measuring the face to determine the location of certain facial points (e.g. the recited feature points), which correspond to the "identified feature points" above. The "set of representative

images" is the video recorded in section I; all faces would have the same general set of features, e.g. eyes, nose, et cetera. The system of Cosatto clearly synthesizes a geometric component, e.g. synthetic video, with specific emphasis on for example the mouth, section V-B (pages 159-161) with other facial parts discussed in section V-D (page 161), which clearly constitutes "generating a geometric component", and the selected image is simply one frame of video wherein the synthesized face is saying something (e.g. see section V-B).)(Moulton clearly teaches that vectors represent the control points of each image and their motion path in the set of archival video footage and that they are tracked between frames [0011, 0059-0061]. See Figure 1 additionally)

-Generating the selected image for the frame from a composite of the set of representative images based on the geometric component. (Clearly, the system of Cosatto generates selected images (as set forth above) from a set of representative images (e.g. see Fig. 6, page 160, where "from a phonetic transcript, parameters of the mouth are calculated ... to obtain an animation script". The "phonetic transcript" is a video clip of an individual saying a phoneme as set forth in sections V-B and V-C on pgs 159-161. Clearly, the images are sampled for the mouth, which would be for example the recited "geometric component". Finally, video – and the image database as recited in section III, pages 153-154 – prima facie teach the use of a "composite" of a set of representative images.")

The section in the Response to Arguments block above concerning the single frame limitation is incorporated by reference.

As stated and detailed above, the Cosatto reference teaches most of the limitations of the above claim. The system of Cosatto performs a specific application – processing of video to generate a database of image components that are then analyzed and used to synthesize output images based on this database and the pre-recorded video. Cosatto does not expressly teach the geometric component is a dimensional vector of feature point positions, though this is implicit (see Figures 3-5).

The system of Moulton performs a similar task, in that it generates new synchronized audio and video of an individual speaking phrases and text that they did not speak before in a realistic manner (see Abstract). The speech and emotions would need to be synchronized, as per the system of Moulton, since this would make speakers more realistic and more believable. The techniques of Moulton for tracking feature points between frames would be ideal for use in the system of Cosatto, which is silent on how the feature points are tracked in the intermediately generated frames per se. However, keeping such data in vector format makes sense, because the plurality of intermediate data generated would be in video format, for example for a video game. Moulton teaches that this technique provides advantages [0020, 0043, 0045, 0057] by allowing the user to associate trajectory paths with transitions and use a standard model for this. The system of Cosatto further uses vector path models for transitions in any case (Figure 6), but Moulton makes such a transition more explicit.

Further, Moulton clearly teaches that standard morphing techniques [0007, 0019, 0024] can be used to perform the transition between phonemes and visemes, with interpolation between positions. Now, one of the primary arguments deployed by

applicant in the Remarks is that Cosatto only finds the closest possibly match to each phoneme pronounced but does not actually **generate** new images. Moulton clearly does to smooth the transitions between captured image photos of various phonemes [0006-0007, 0009-0013, etc], where this would clearly smooth the transitions between the closest matches of the Cosatto system and would not change the principle of operation.

However, keeping such data in vector format makes sense, because the plurality of intermediate data generated would be in video format, for example for a video game. Moulton teaches that this technique provides advantages [0020, 0043, 0045, 0057] by allowing the user to associate trajectory paths with transitions and use a standard model for this. Cosatto utilizes three-dimensional models [0064], where perspective projection techniques are used to render three-dimensional models and scenes (such as a 3D model of a face) to the two-dimensional screen, as does Moulton (inherently, by generating video). Therefore, for at least the above reasons, it would have been obvious to one of ordinary skill in the art at the time the invention was made to modify Cosatto to use vectors to store feature point paths and morphing as specified above.

As to claim 2, Cosatto clearly states that certain feature points are measured initially – see page 154, where in Fig. 1 and section IV-A, the measurement of a few feature points are done. This allows the system to track only those points in video, where however each of those points is recalculated per frame; these points clearly constitute the recited "plurality of values" with at least one value associated with at least each representative image. Generating the recited synthesized image (as taught on

pages 159-161), especially for the mouth region, clearly requires using the plurality of values to composite the video frames to generate the image database and to composite the synthesized image so that it has photorealism.

As to claim 3, the claim requires that the synthesized image and the representative or source images be a plurality of subregions that are near each other. Clearly a face consists of regions that are proximate to each other, as shown in Cosatto in Fig. 1 for example, and discussed in detail in the design of the database for facial parts (page 154, sections III-A→III-D). Therefore, as the images are analyzed and put into the database, and the geometric components (e.g. the lips and mouth cited in section V-B) generated, and the final output images synthesized (section V in its entirety, sections V-B through V-D, pages 159-161, as well as Figure 8) all of these use portions of the face subdivided into regions or quads.

As to claim 4, Cosatto on pg. 161, section V-E, clearly states that the base bitmap of the face is put into a frame buffer, then the bitmaps of the facial parts are projected onto it, and finally that such bitmaps are blended ("gradual blending or "feathering" masks").

As to claim 5, Cosatto teaches it for the same reason as above, namely that on page 161, on the left side, immediately below Fig. 8, it teaches that regions are blended and further that these are facial regions – obviously, regions of the face (except perhaps the mouth) will have common texture – in that they will all be covered in skin, and thusly alpha blending is used as set forth there, so those boundaries would not have discontinuities in texture.

As to claim 6, that would be a trivially obvious variant, given that the system of Cosatto is intended primarily for synthesizing the mouth region to create natural-appearing speech, there would obviously be more samples of the mouth region than of other regions, particularly in the database of parts, as that is derived from all the video-recorded phonemes. Therefore, either Cosatto implicitly teaches it or it is a trivially obvious variant, and it would be obvious to modify for the reasons set forth immediately above, and on page 59, section H, it is stated the mouth database is larger than those of other features and an absolute size provided.

As to claim 7, Cosatto on page 155 teaches specific items wherein section II-A teaches that the 3-D head pose can be recovered from 2-D images and at the bottom right of the page it states that 2-D head models provide acceptable ranges for size. Further, the system uses video cameras to capture images for analysis purposes, and it is prima facie obvious that the feature points are found in 2-D images as recited in the upper left section of the page, e.g. section A.

As to claim 8, it is a duplicate of claim 3 and the rejection to claim 3 is incorporated herein by reference in its entirety.

As to claim 9, it is a duplicate of claim 4; see that rejection.

As to claim 10, this is a trivially obvious variant of claim 7, wherein Cosatto in section 2 (pages 152-154, emphasis on page 153) states that three-dimensional images using 3-D scanners are common in the art and in prior work. As further discussed in section IV-A on page 154, feature points on the face are measured in 3-D. Therefore, it would be obvious that the feature points could be on a three-dimensional image and it

would be obvious to modify the system of Cosatto to use three-dimensional images for the reasons set forth above.

As to claim 11, it is a duplicate of claim 3 and the rejection to that claim is herein incorporated by reference – only the one reference is utilized.

As to claim 12, it is a duplicate of claim 4; see that rejection.

As to claim 13, Cosatto teaches in section A on page 155 that "Knowing the position of a few points in the face allows to recover the 3-D head pose from 2-D images", where this clearly justifies that examiner's contention that the a few key feature points are used to extract the position of other feature points, see for example section VI-D on page 157. Section V-B on pages 159-160 clearly teaches how knowledge of a few points allows synthesis of a great many essential feature points on the mouth, which is the key feature.

As to claim 14, Cosatto teaches that obviously feature points are grouped in sets by different regions of the face – see page 154, sections III-1 through III-4 and Fig. 1 or of the synthesized image – see page 161, sections D and E. Finding the position of one feature point on for example the mouth (see section V-B and V-C, particularly page 160) allows the calculation of the shift in other portions of the images, e.g. where the changes in position between one frame and another of the synthesized image are minimized to get more natural appearance (e.g. Figure 7 on page 160), which prima facie tracks change in position of feature points within the mouth region so as to be able to calculate the path that involves the least change in position for Viterbi optimization,

and the details on feature point location and tracking are found in sections III-D and III-E, particularly section III-D. Thusly, Cosatto teaches all the limitations.

As to claim 24, on page 153 Cosatto teaches that in the left hand column that prior work in the field has dealt with aligning images to a reference image; thusly it would be obvious that the system of Cosatto could perform the recited limitations by aligning the generated facial image with a reference image, since that would allow deviations due to motion from corrupting the base image that the system was working with, given that the system of Cosatto puts the base shape into the frame buffer and then projects the other facial regions onto it, with the use of a reference image being an obvious variant on the "base image" require to do so – this is discussed in the rejections to the claims above, and motivation is taken from claim 1 above.

As to claim 25, firstly, the rejection to claim 1 is herein incorporated by reference in its entirety. That rejection incorporates the limitations of having and accessing a database of various portions of a face and various captures of a full face with corresponding matching feature point. Secondly, the rejection to claim 14 is incorporated by reference in its entirety. That rejection teaches the limitation of determining a feature point from a change in position of another feature point based on a change in the selected feature point and the existence of a database (page 153) containing facial information (also, the hierarchy of that database is in page 154, section 1). Next, Moulton clearly does teach a database of various stored representations (e.g. video and phonemes), where the actor clearly has the same feature points in each frame, since the actor is speaking or the like, as does Cosatto. Clearly, Moulton does in

fact track the location of the feature points as specified in between each frame, where the feature points being tracked constitute selected feature points, and clearly as discussed in the rejection to claim 1 the new frame is rendered with the feature points having changed position, since speech is being simulated and/or the transitions between various phonemes and speaking frames.

Claims 15-23 and 25-43 are rejected under 35 U.S.C. 103(a) as being unpatentable over Cosatto and Moulton as applied to claim 14 above, and further in view of Chai et al (Chai et al. "Vision-based control of 3D animation".)

As to claim 15, Cosatto and Moulton do not expressly teach the limitation of using PCA to track position. Chai teaches the use of PCA on pages 200-201 for example with emphasis on sections 4.2 and 4.3, where it is taught that using PCA, the motion frames are broken down into linear subspaces and motion is tracked in that way. On page 200, section 4.3 it clearly discloses that a database of motion is kept, which would be similar to the database of images in Cosatto and Moulton. The database of motion would be with respect to each linear subspace, which obviously could be the different facial regions of Cosatto and Moulton – that is, the positional changes in motion of the images in the database of Cosatto and Moulton for facial regions could be found using the PCA techniques of Chai. Therefore, It would have been obvious to one having ordinary skill in the art at the time the invention was made to combine the PCA of Chai with the motion tracking and splitting of the face into different regions of Cosatto and Moulton for the reasons set forth above, as using PCA allows faster computation

times for motion detection and improves temporal coherency (pg. 200- section 4.4 for example).

As to claim 16, it would have been obvious that given that Cosatto and Moulton tracked motion using overall feature points on the face (e.g. section III-A page 155 or section D page 161) and that Cosatto and Moulton also tracked feature points within the mouth subset in order to assure more natural appearing features as the difference between each pose was minimized via Viterbi optimization on page 160, Figure 7. Obviously, overall changes in head position would tracked via the main feature points and determining the necessarily positional changes in the mouth (besides those necessitation by normal motion of talking) would be based on the positional changes in the larger set of feature points on the face itself, e.g. any necessary translational or rotational movement of the overall head for example. Since only the primary reference is utilized, no separate motivation or combination is required and that from the rejection to the parent claim is herein incorporated by reference.

As to claim 17, the system of Cosatto and Moulton has a hierarchical database structure of feature parts, see for example section III, page 154, items 1-3, particularly Item I, titled "Hierarchy of Parts." Since only the primary reference is utilized, no separate motivation or combination is required and that from the rejection to the parent claim is herein incorporated by reference.

As to claim 18, the system of Cosatto and Moulton does not expressly teach this limitation, insofar as it teaches tracking feature points of the user when the data for the initial sets is recorded, but it does not expressly perform the recited details, although it

does monitor feature points of a user. The system of Chai performs the recited limitations, in that it consists of a video camera that monitors the face of a user and generates an image of an avatar making similar facial movements, see for example Fig. 1 on page 193, the caption specifies that users act out the motion in front of a single-view camera, and that the avatars have controlled facial movements similar to those of the user with texture mapped models (see section 1, left side of page 194, and Figure 2 on page 195, and the captions on it). The system of Chai further tracks feature points of the user (section 2.1, page 196) on the face and moves the avatar as the user moves (see section 1.2 on page 196, where motion data and head motion are separated from facial deformations and then both are applied to the avatar in separate passes). Obviously, the generated avatars of Chai (Figs. 1 and 2 for example) have separate components of the face, or it would be obvious to use the separate components of Cosatto and Moulton for the face, and to utilize the motion tracking and facial deformation techniques of Chai described above. It would have been obvious to one having ordinary skill in the art at the time the invention was made to combine the systems of Cosatto and Moulton and Chai, since Chai would allow any user to control the facial expressions of an avatar in addition to overlaying audio text and simulating real speech – the facial techniques would allow better synchronization of voice and facial movements in for example the avatars, and would allow even an unskilled user to adequately control facial motions (see section 1, pages 193-194).

As to claim 19, Cosatto and Moulton teaches in section A on page 155 that "Knowing the position of a few points in the face allows to recover the 3-D head pose

from 2-D images", where this clearly justifies that examiner's contention that the a few key feature points are used to extract the position of other feature points, see for example section VI-D on page 157. Section V-B on pages 159-160 clearly teaches how knowledge of a few points allows synthesis of a great many essential feature points on the mouth, which is the key feature. Since only the primary reference is utilized, no separate motivation or combination is required and that from the rejection to the parent claim is herein incorporated by reference.

As to claim 20, reference Cosatto and Moulton does not expressly teach this limitation, insofar as it does teach rendering an image of a speaking human being with the identified feature points on it (see for example Fig. 8, and facial locations are tracked by feature points as illustrated by Fig. 4, where the control points are noted. However, Chai teaches on page 196 in the "initialization" section that the user can select the control points, for which it would be an obvious modification to allow the user to control the movement of a feature point. Also, since the system of Chai (for example, see caption on Fig. 1 on the first page) teaches that the avatar moves in response to user facial and head movements, this also constitutes "receiving information indicative of a user moving a feature point". It would have been obvious to one having ordinary skill in the art at the time the invention was made to combine the systems of Cosatto and Moulton and Chai, since Chai would allow any user to control the facial expressions of an avatar in addition to overlaying audio text and simulating real speech – the facial techniques would allow better synchronization of voice and facial movements in for

example the avatars, and would allow even an unskilled user to adequately control facial motions (see section 1, pages 193-194).

As to claim 21, this claim is a substantial duplicate of claim 16; the rejection to that claim is herein incorporated by reference in its entirety, along with motivation and combination.

As to claims 22 and 23, Cosatto and Moulton does not expressly teach this limitation, whilst Chai teaches in Fig. 1 on page 193 that the user can control or select the facial expression by making the desired expression on their own face, e.g. two separate facial expressions are shown in the leftmost column, and in the rightmost column the avatars are shown depicting those facial expressions. Motivation and combination is incorporated by reference from claim 20 above.

As to claim 25, firstly, the rejection to claim 1 is herein incorporated by reference in its entirety. That rejection incorporates the limitations of having and accessing a database of various portions of a face and various captures of a full face with corresponding matching feature point. Secondly, the rejection to claim 14 is incorporated by reference in its entirety. That rejection teaches the limitation of determining a feature point from a change in position of another feature point based on a change in the selected feature point and the existence of a database (page 153) containing facial information (also, the hierarchy of that database is in page 154, section 1). That is, Cosatto and Moulton teaches those limitations and Chai teaches a motion database on page 194 on the left side of the page in section 1 and in the caption to Figure 2 on page 195, Chai teaches that the motion database can be used to

synthesize expressions. Obviously, displaying the avatars on the screen as shown in Figure 1 of Chai clearly constitutes rendering (see for example page 202, the reference to the terms "rendering cost"). Finally, it would be trivially obvious that since the user controls the motion and deformation of the avatar by means of controlling their own head position and facial expressions, and that Chai (e.g. section 1.2, page 196) clearly teaches separating head motion and facial expressions such that Chai's system would obviously render a new expression based on any change of feature points, with the choice of two being a trivially obvious variant. It would have been obvious to one having ordinary skill in the art at the time the invention was made to combine the systems of Cosatto and Moulton and Chai, since Chai would allow any user to control the facial expressions of an avatar in addition to overlaying audio text and simulating real speech – the facial techniques would allow better synchronization of voice and facial movements in for example the avatars, and would allow even an unskilled user to adequately control facial motions (see section 1, pages 193-194).

As to claim 26, this claim is essentially a duplicate of claim 14, with the difference that Cosatto and Moulton teaches that the feature points are grouped in sets according to the region of the face, e.g. the hierarchical database shown on page 154, and the rest of the limitations are taught in the rejection to claim 14, which is herein incorporated by reference in its entirety. Motivation and combination are taken from claim 25 above.

As to claim 27, this claim is a substantial duplicate of claim 15, with the only difference that the database of representative images of Cosatto and Moulton is substituted for the motion database of Chai. Chai teaches a motion database on page

194 on the left side of the page in section 1 and in the caption to Figure 2 on page 195, Chai teaches that the motion database can be used to synthesize expressions. The rest of the limitations are taught in the rejection to claim 15, which is herein incorporated by reference in its entirety; motivation and combination are taken from claim 25, which is the parent claim.

As to claim 28, this claim is a substantial duplicate of claim 16, with that rejection herein incorporated by reference; motivation and combination is from claim 25 above.

As to claim 29, this claim is a substantial duplicate of claim 17, with that rejection herein incorporated by reference; motivation and combination is from claim 25 above.

As to claim 30, this claim is a substantial duplicate of claim 3, with the rejection to that claim incorporated by reference in its entirety. The only difference is that the database of representative images of Cosatto and Moulton used in the rejection to claim 3 is replaced with the motion database for generating expressions of Chai for the reasons set forth in the rejections to claims 25 and particularly claim 27 above, which rejections are also herein incorporated by reference. The motivation and combination is taken from claim 25.

As to claim 31, this claim is a substantial duplicate of claim 4, the rejection to which is incorporated herein by reference. Since only the primary reference is utilized, no separate motivation or combination is required and that from the rejection to the parent claim is herein incorporated by reference.

As to claim 32, this claim is a substantial duplicate of claim 5, the rejection to which is incorporated herein by reference. Since only the primary reference is utilized,

no separate motivation or combination is required and that from the rejection to the parent claim is herein incorporated by reference.

As to claim 33, this claim is a substantial duplicate of claim 6, the rejection to which is incorporated herein by reference. Since only the primary reference is utilized, no separate motivation or combination is required and that from the rejection to the parent claim is herein incorporated by reference.

As to claim 34, Cosatto and Moulton does not expressly teach this limitation, whilst the system of Chai performs the recited limitations, in that it consists of a video camera that monitors the face of a user and generates an image of an avatar making similar facial movements, see for example Fig. 1 on page 193, the caption specifies that users act out the motion in front of a single-view camera, and that the avatars have controlled facial movements similar to those of the user with texture mapped models (see section 1, left side of page 194, and Figure 2 on page 195, and the captions on it). Motivation and combination are taken from the parent claim, e.g. 25 and herein incorporated by reference.

As to claim 35, this claim is merely claim 25 with the limitations of claim 20 added to it. The rejections to both claims are herein incorporated by reference in their entirety, and the motivation is taken from claim 25.

As to claim 36, this claim is a substantial duplicate of claim 26, with that rejection herein incorporated by reference; motivation and combination is from claim 25 above.

As to claim 37, this claim is a substantial duplicate of claim 27, with that rejection herein incorporated by reference; motivation and combination is from claim 25 above.

As to claim 38, this claim is a substantial duplicate of claim 28, with that rejection herein incorporated by reference; motivation and combination is from claim 25 above.

As to claim 39, this claim is a substantial duplicate of claim 29, with that rejection herein incorporated by reference; motivation and combination is from claim 25 above.

As to claim 40, this claim is a substantial duplicate of claim 30, with that rejection herein incorporated by reference; motivation and combination is from claim 25 above.

As to claim 41, this claim is a substantial duplicate of claim 31, with that rejection herein incorporated by reference; motivation and combination is from claim 25 above.

As to claim 42, this claim is a substantial duplicate of claim 32, with that rejection herein incorporated by reference; motivation and combination is from claim 25 above.

As to claim 43, this claim is a substantial duplicate of claim 33, with that rejection herein incorporated by reference; motivation and combination is from claim 25 above.

Conclusion

The prior art made of record and not relied upon is considered pertinent to applicant's disclosure: US PGPub 2001/0031081 to Quan, which discloses a digital mirror that performs the recited limitations. Applicant is put on notice that should the previous rejections for any reason be vacated, set aside, reversed, and/or withdrawn, new grounds of rejection will be added using the Quan reference.

Applicant's amendment necessitated the new ground(s) of rejection presented in this Office action. Accordingly, **THIS ACTION IS MADE FINAL**. See MPEP § 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the date of this final action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Eric Woods whose telephone number is 571-272-7775. The examiner can normally be reached on M-F 7:30-5:00.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Michael Razavi can be reached on 571-272-7664. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

Eric Woods

December 28, 2005

ALMIS R. JANKUS
PRIMARY EXAMINER